Chromatin structure and gene expression

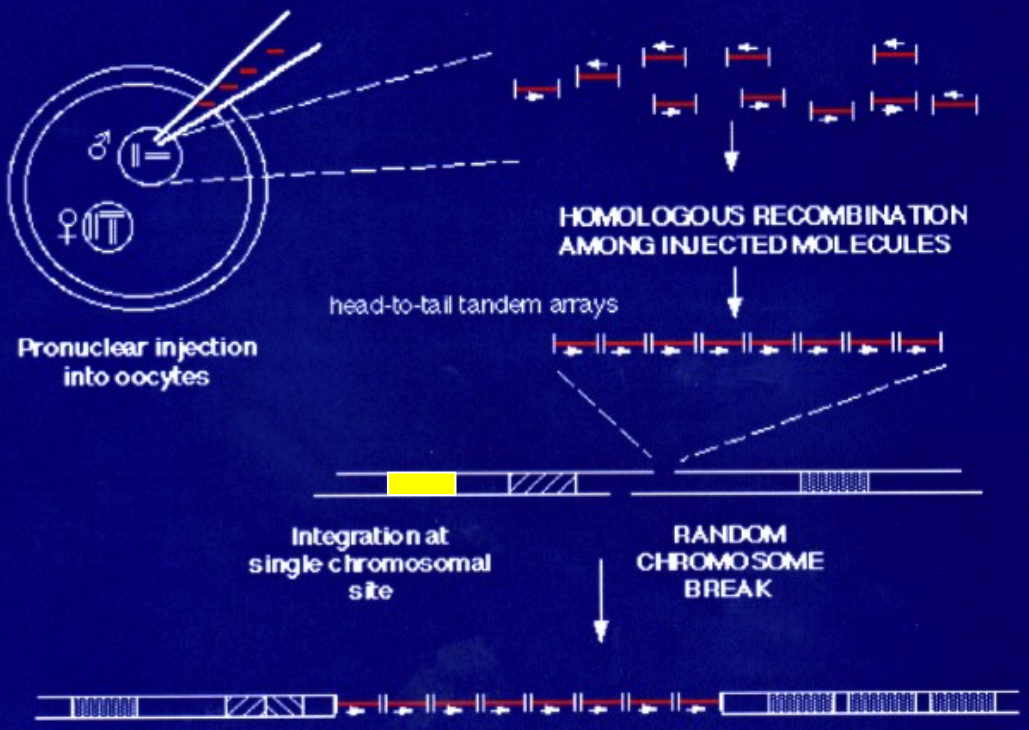
II. ***Long range regulation of gene transcription***

Tuesday, 3rd November, 2015

Andy Bannister

ab149@cam.ac.uk

TRANSGENE INTEGRATES RANDOMLY IN THE GENOME

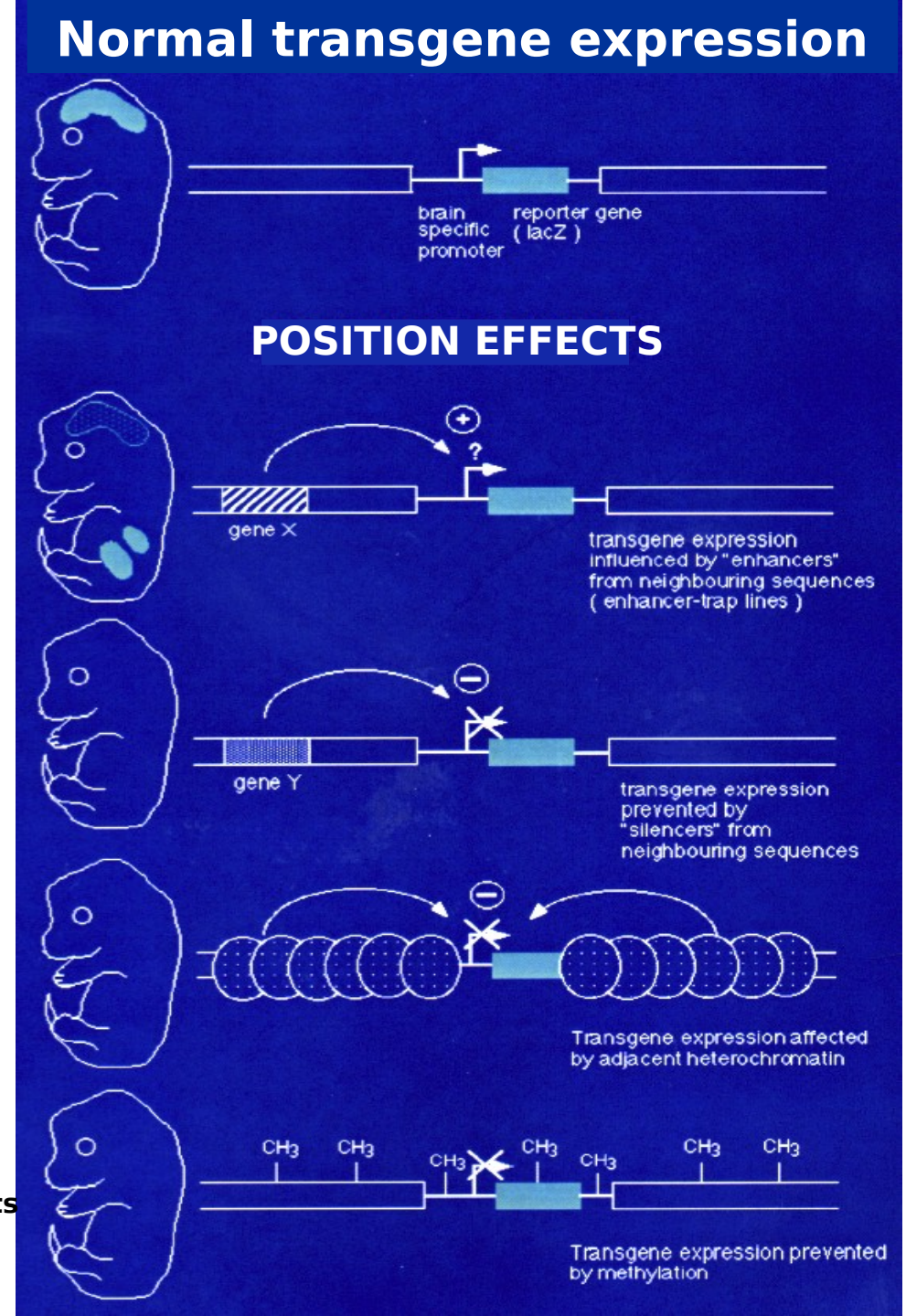


INSERTIONAL EFFECTS

Variable number of integrated genes

(Potential) Problems:

- At the insertion site (host sequences)
  - duplications
  - rearrangements
  - deletions
  - translocation
- Insertional mutation (~ 10%) mostly resessive
  - lethality (dominant)
  - sterility (insertion site important for reproduction)



This example is a single integrated gene

Enhancers

Silencers

Heterochromati

DNA methylatio

Position-effect variegation (PEV) is a variegation caused by the inactivation of a gene in some cells through its abnormal juxtaposition with heterochromatin.

Factors Affecting Transgene Expression

These ‘factors’ are defined **functionally**

1. Enhancers

2. Silencers

3. Heterochromatin
   - -Locus Control Regions (LCRs)
   - -Matrix Attachment Sites (MARs)
   - -Insulators / Boundary Elements
   - -RNAi

4. DNA Methylation

1. Enhancer elements

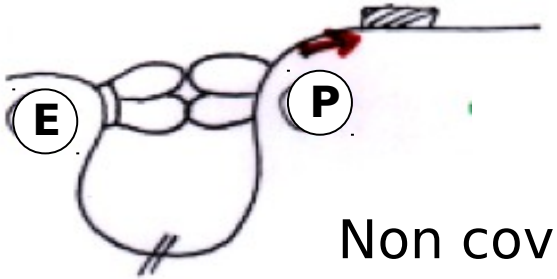
- DNA element that enhances transcription of reporter gene in **transient transfection** *a*ssay
- DNAse hypersensitive sites (HSs) -- Accessible !
  - **transcription factor binding site(s)**

Enhancers also bind factors like the p300 coactivator which is a HAT

(Not true chromatin)

- **orientation independe**
- **upstream/downstrean**
- **proximal/distal**

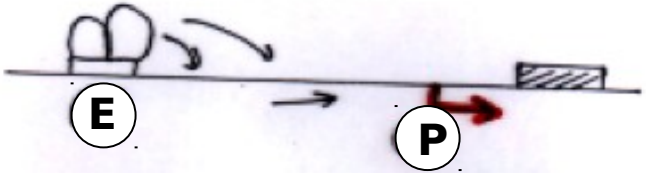
Possible mechanism of function:
- Looping / spreading
- Tracking mechanism



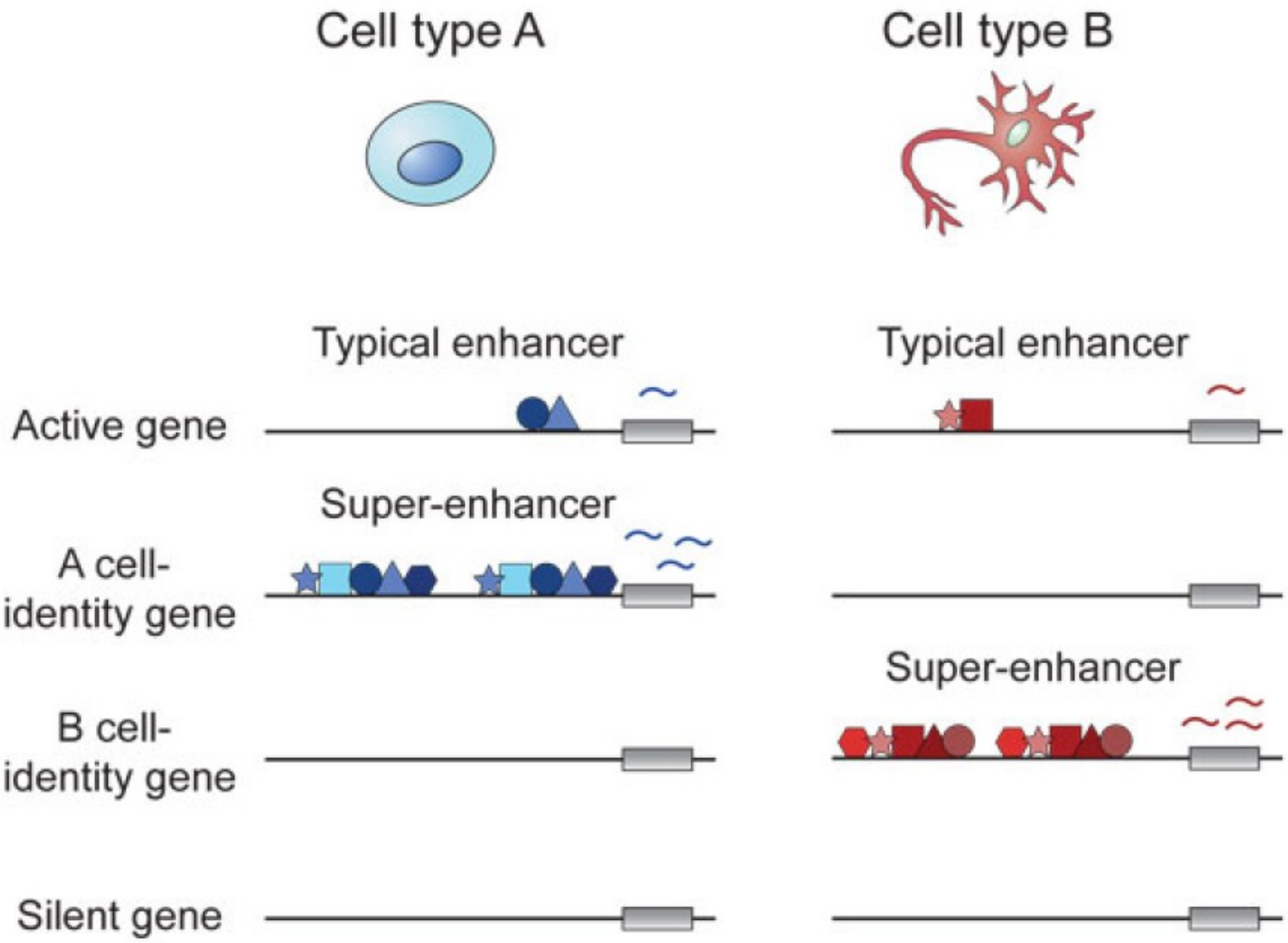
Marked by specific histone modifications such as H3K4me1 and H3K27ac and 'enhancer' RNAs (eRNAs)

Non covalent linkage reqd

Expressed at high but variable levels in transgenic mice when integrated near heterochromatin, i.e. it's subject to



There are Typical Enhancers and Super-Enhancers

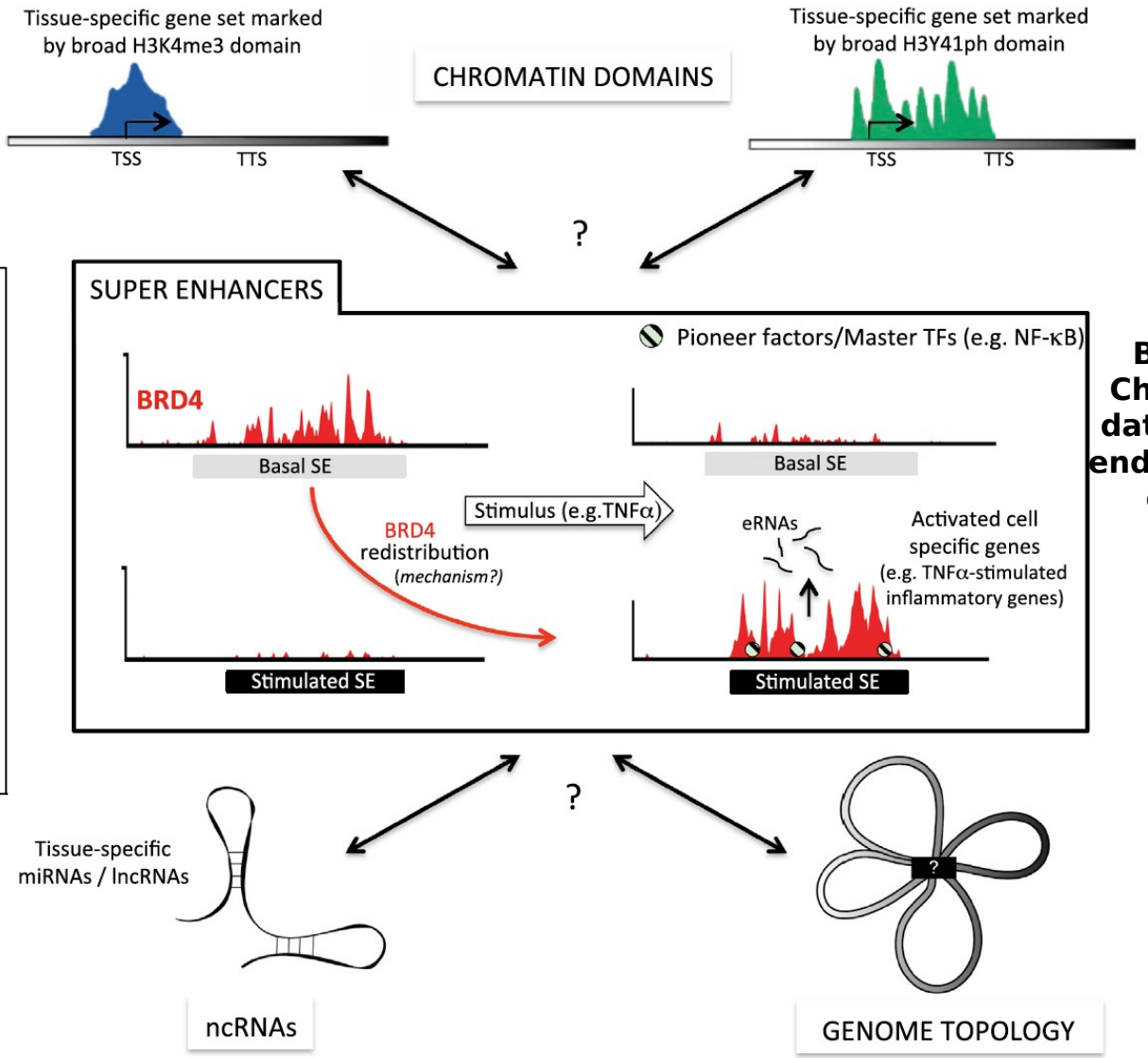


SEs have very high levels of H3K4me1, H3K27ac and BRD4 (a bromodomain p

Super-Enhancers and their interplay with other mechanisms controlling gene expressio

Super-enhancers (SEs) are linked to cell fate.

The dynamic nature of SEs can be seen when **endothelial cells** are activated by TNF-α . The formation of *de novo* SEs and concomitant decommissioning of pre-existing enhancers drives rapid inflammatory transcriptional responses and involves NF-**k**B-dependent redeployment of **BRD4**.



2. Silencers

CLASSICAL SILENCER:

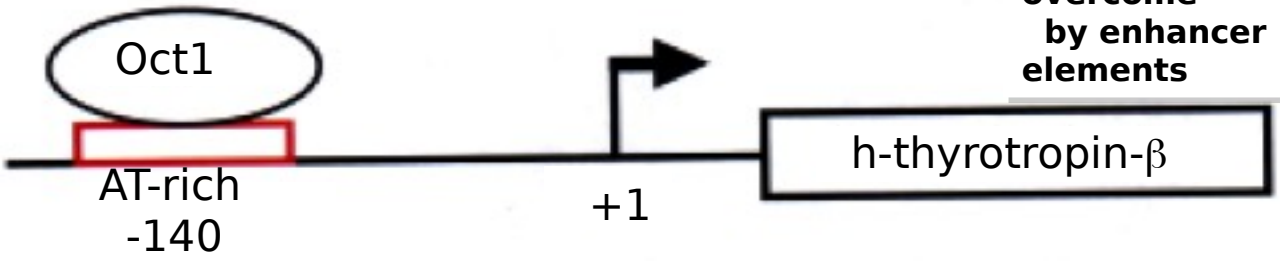
Activity associated with relatively SHORT DNA elements
Antithesis of enhancers
- E.g. human thyrotropin-β

ALSO, Longer more complex DNA elements
eg PcG silencing
yeast telomeres

Involves formation of heterochromatin

over many kilobases

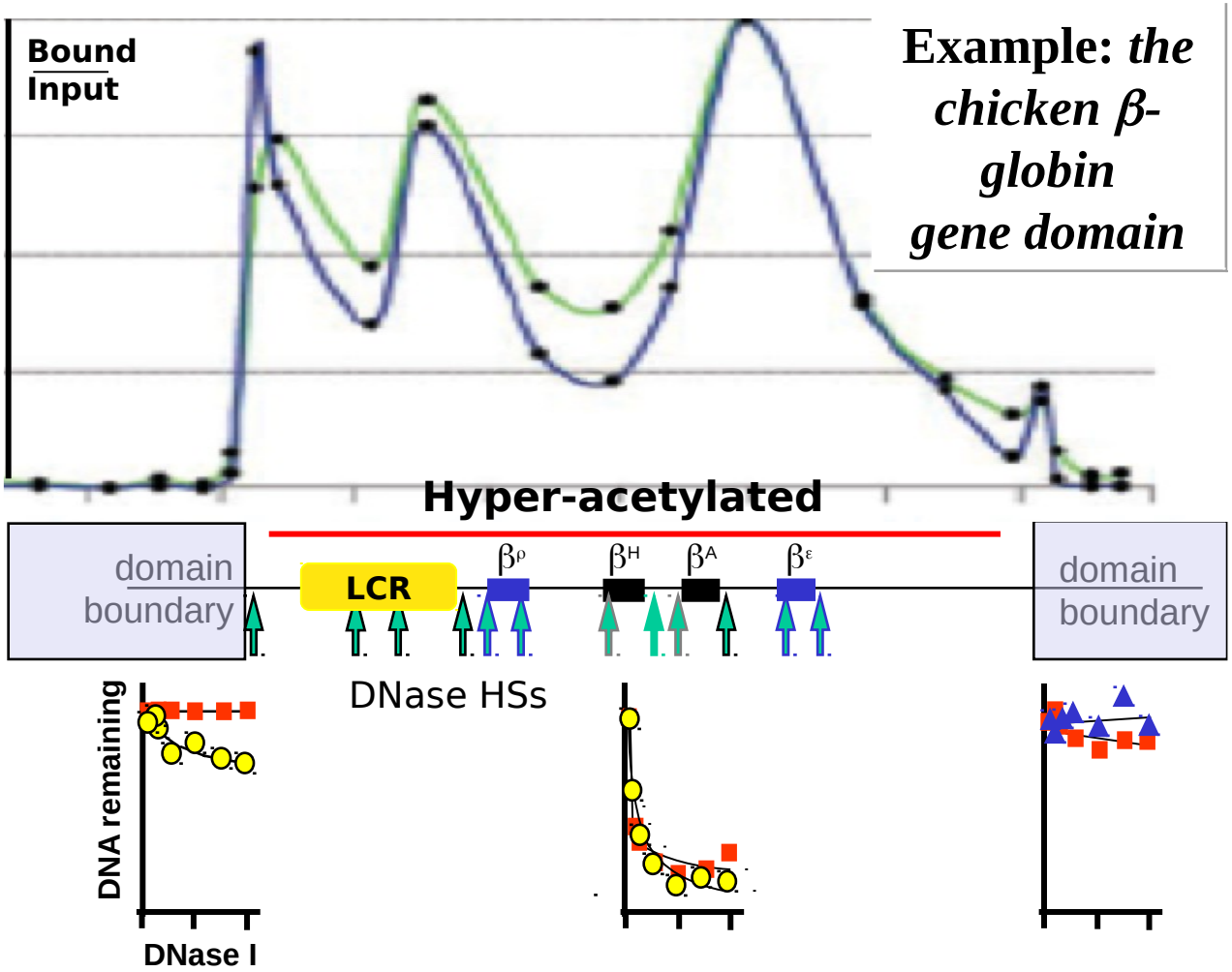
CLASSICAL SILENCERS
Eg human thyrotropin-β gene

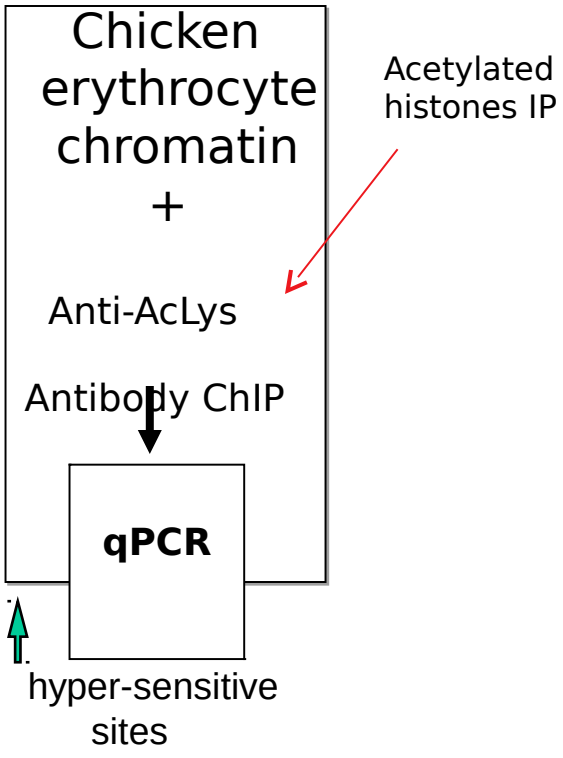


Silencer can be overcome by enhancer elements

- Orientation/position independent - TF binding site(s) - DNase HSSs
- Expression restricted to thyrotrophs
- Above construct subject to position effects (PEV) in transgenic mice when integrated near heterochromatin

3. Locus Control Regions





General message; Lysine acetylation in histones correlates with gene activity

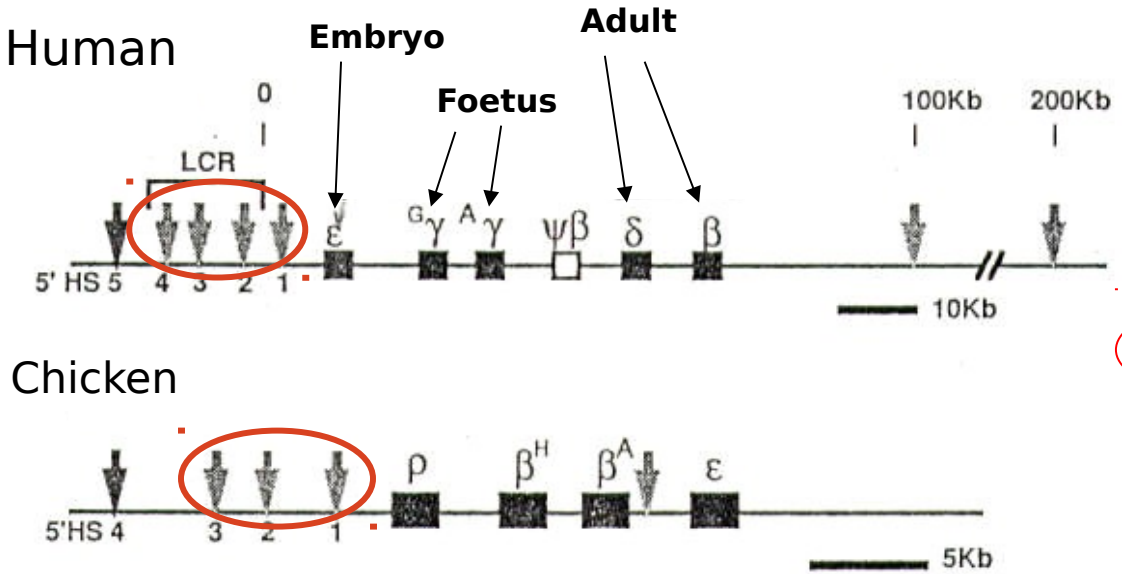
Acetylation also correlates with di-/tri-methylation of K4 in histone H3

These histone marks define a region of transcriptional 'competence'

3. Locus Control Regions

- defined in β-globin locus, 5' cluster of DNase I HSs

DNase HSs



Evolutionarily conserv

* -from mini gene constructs: LCR required for high level, copy number dependent (position indepe gene expression
-LCR required for tissue specific and developmental regulation

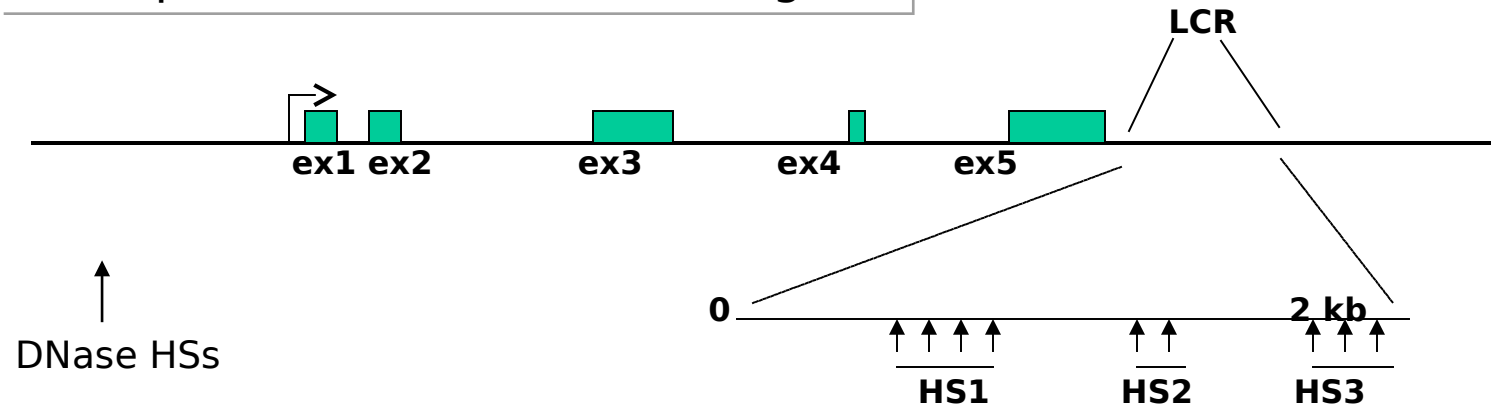
HSs - what are they?

- individual HS weaker effect than full length LCR
- HSs have core region of 200-300bp --- binding sites for multiple TFs
- HS2 acts as a strong E, but this not true for all HSs --- difference to E elements is not clear

3. LCRs (contd…..)

hCD2 is a cell adhesion molecule expressed on the cell surface

Example 2: LCR in human CD2 gene



-required for high level, position independent (copy number dependent) gene expression

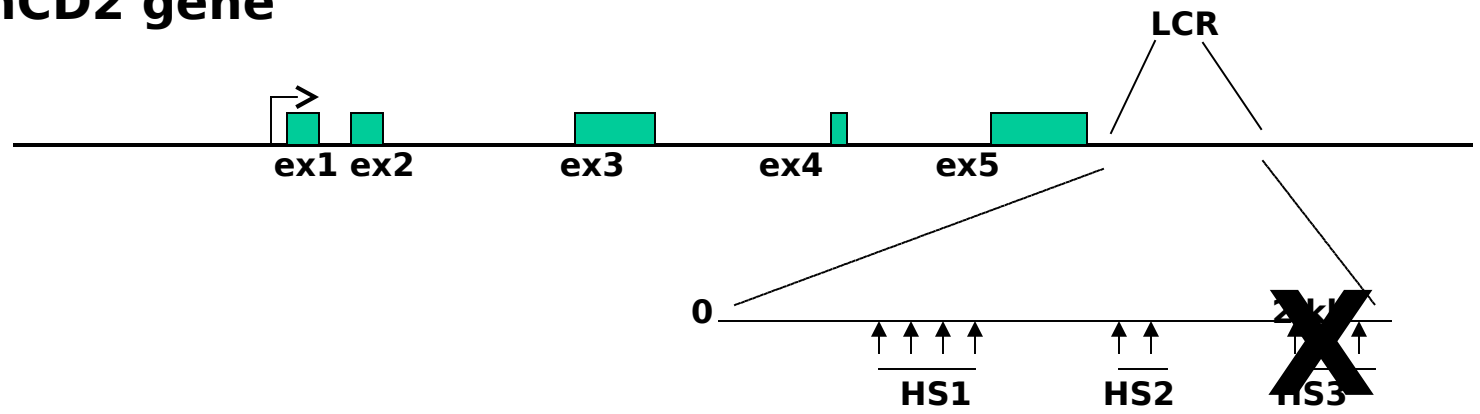
- deletion of HS region 3 in LCR leads to PEV (position effect variegation) when integrated near heterochromatin, activation is a **random** event

- HS region 3 is required to maintain an open chromatin configuration

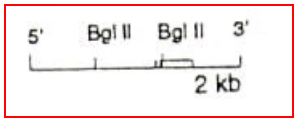
LCRs can open chromatin

hCD2 is a cell adhesion molecule expressed on the cell surface

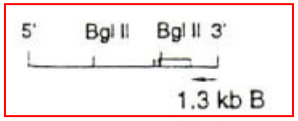
hCD2 gene



INTEGRATED transgene



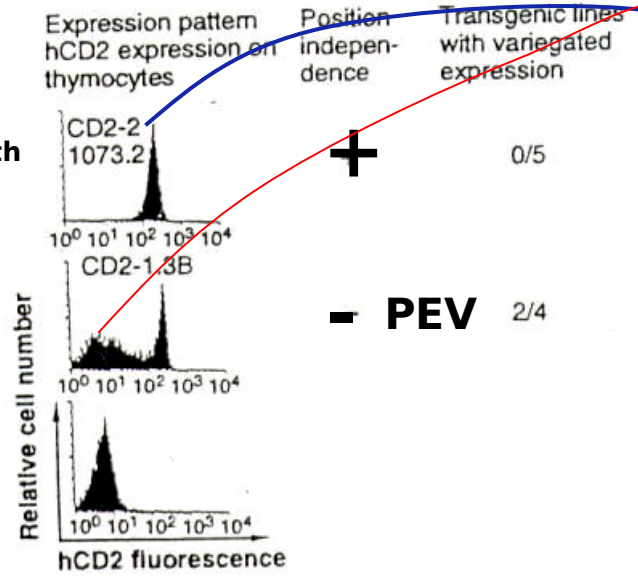
full length hCD2



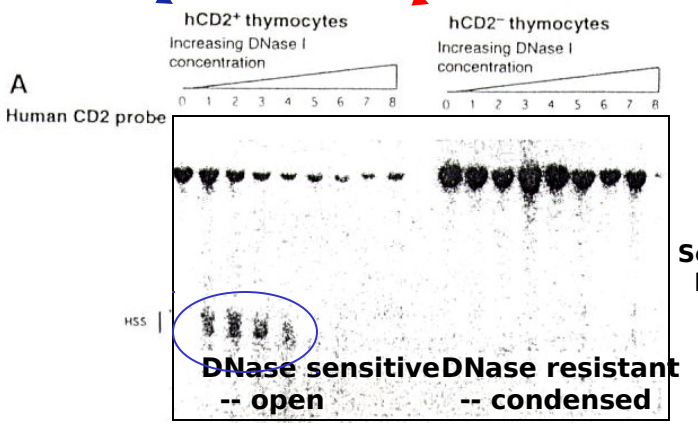
hCD2 ΔHS3

Nontransgenic
WT cells

no hCD2

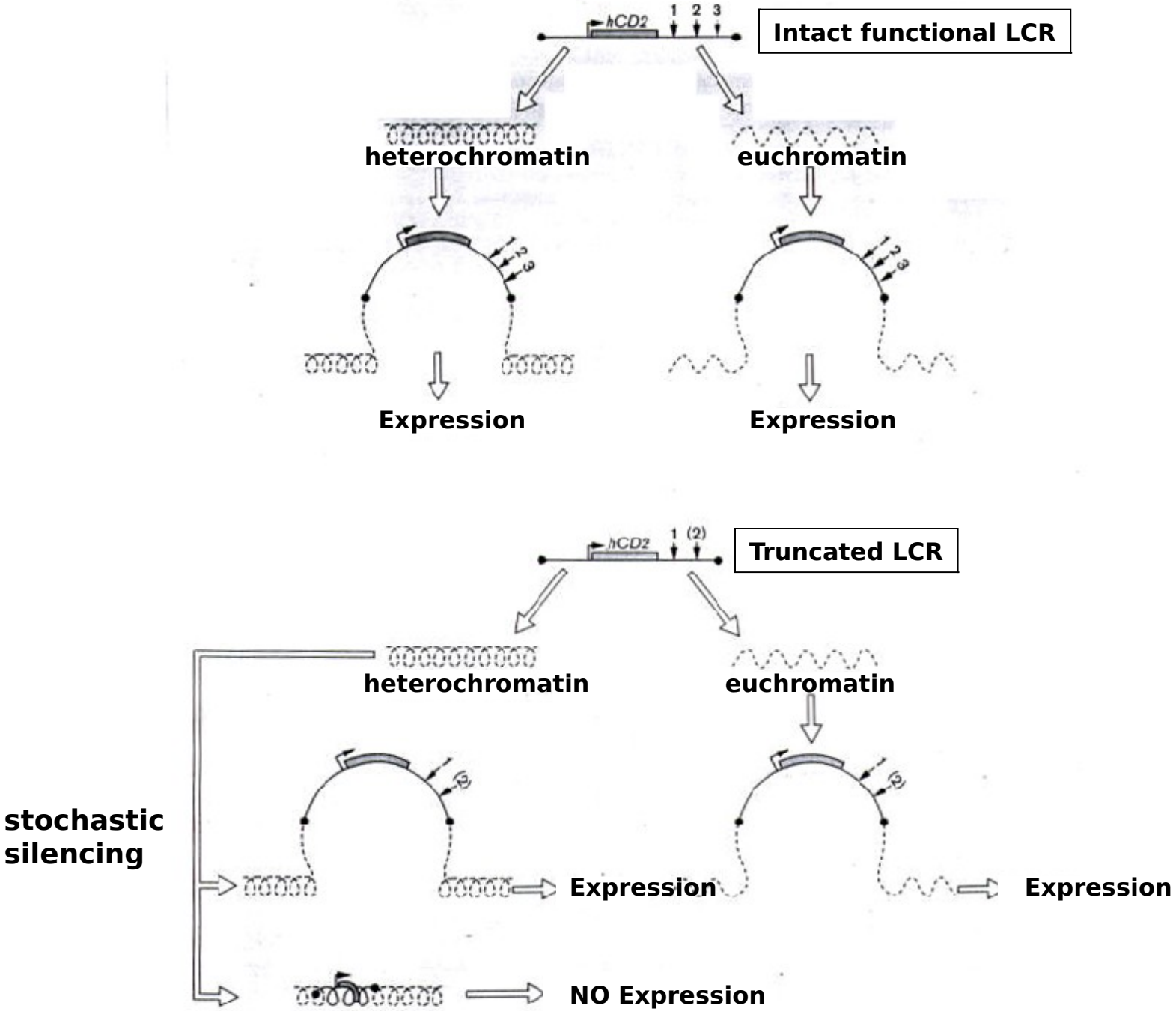


Expression determined by FACS analysis



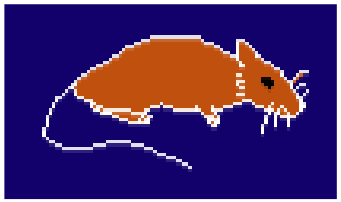
Southern blot

So, cell lines expressing CD2 maintain the gene in an accessible environment but non-expressing cells do not


Intact functional LCR
hCD2
1 2 3
heterochromatin
euchromatin
2 3
Expression
Expression
Truncated LCR
hCD2
1 (2)
heterochromatin
euchromatin
(2)
stochastic silencing
Expression
Expression
NO Expression

Molecular basis of albino phenotype

(tyrosinase gene)



5’ 3’

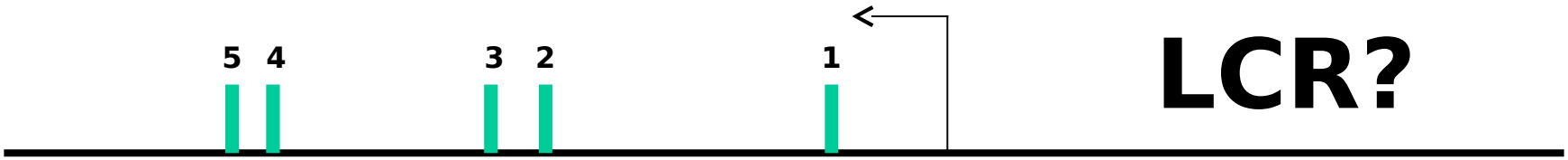
Wild type **... GGA AAC TGT AAG TTT ...**

Gly Asn Cys^{85} Lys Phe

Albino **... GGA AAC TGT AAG TTT ...**

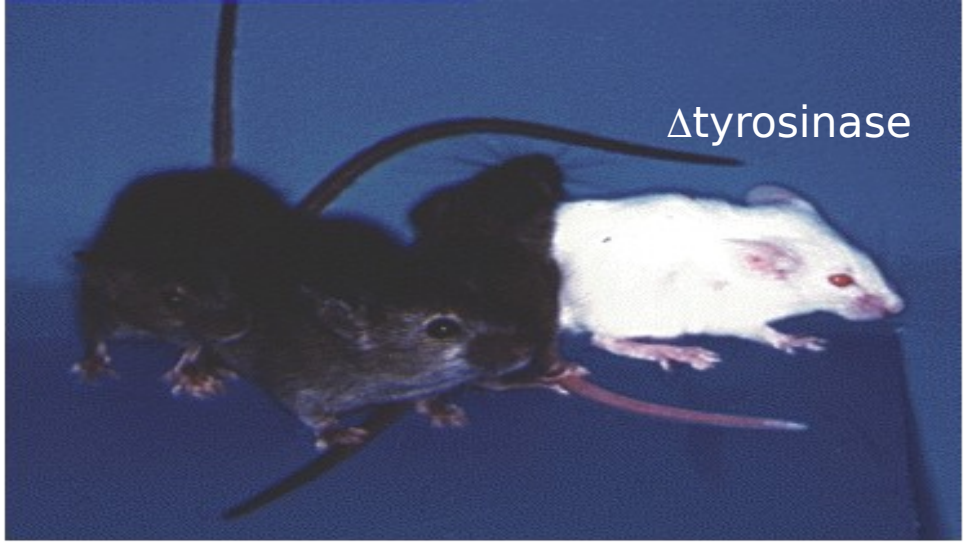
Gly Asn Ser^{85} Lys Phe

Does the tyrosinase gene contain an LCR?



In transgenic mice, does it express in a
copy number dependent (position independent) manner?

WT



Gene copies

0 **0** **1** **2 WT**

Even expression: non-mosaic

4. Matrix or Scaffold Attachment Regions (MARs/SARs)

- no DNAse hypersensitive sites

- digest DNA, purify nuclear scaffold and isolate pieces of DNA that remain associated with scaffold = SARs

- AT-rich sequence {(AT)n(A)mX}5-6

- dual structural and functional role

- Major scaffold attachment protein: Sc1 identified as topoisomerase II --- supercoiling

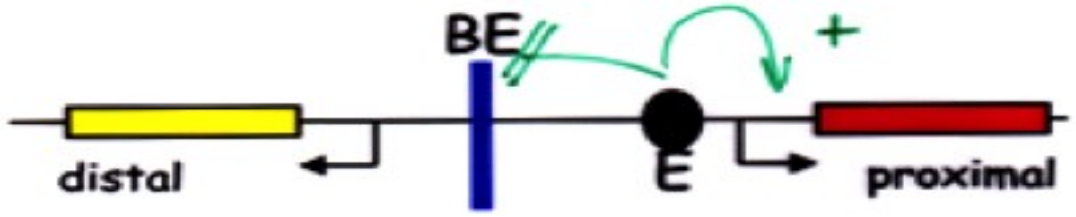
5. Boundary Elements/Insulators

e.g. - ICR at imprinted genes
- Fab elements of Dros. Abd-

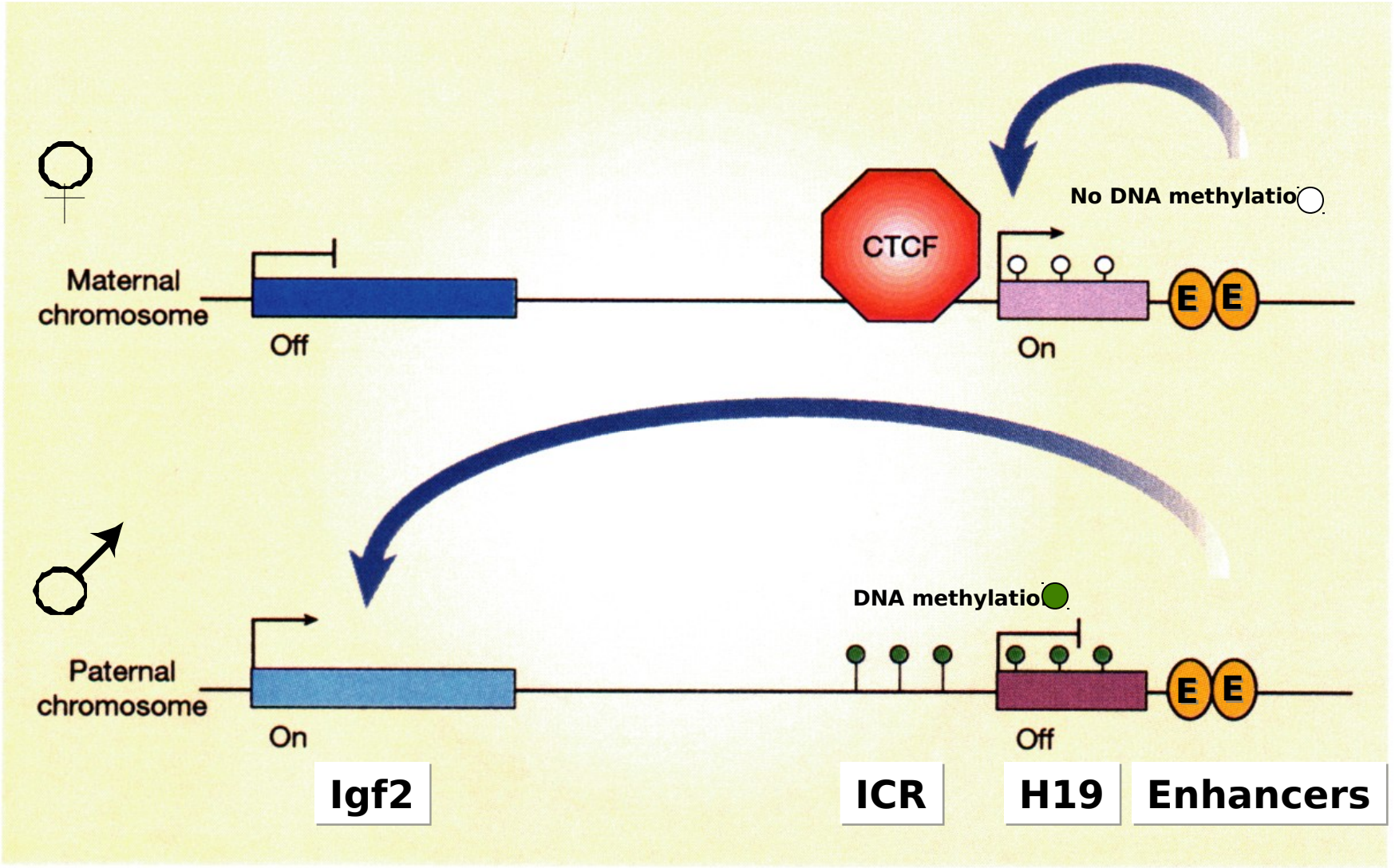
Common features of insulators

1 - insulate genes from position effects - protect from encroachment of heterochromatin

2 - insulate genes from enhancers, no effect on basal activity



CTCF is involved in enhancer blocking at imprinted genes

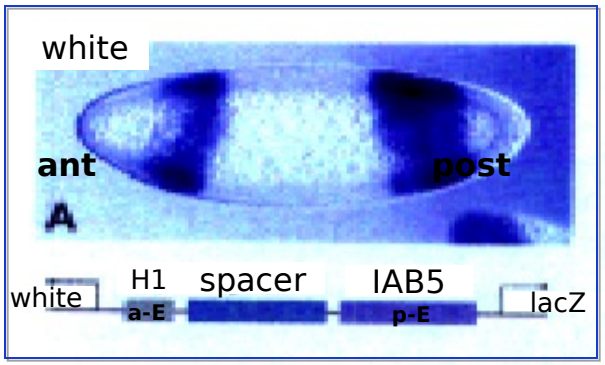


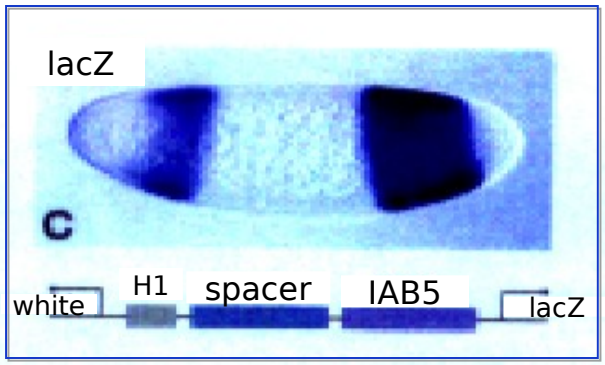
Uncertain whether enhancer blocking is a major function of CTCF at other site

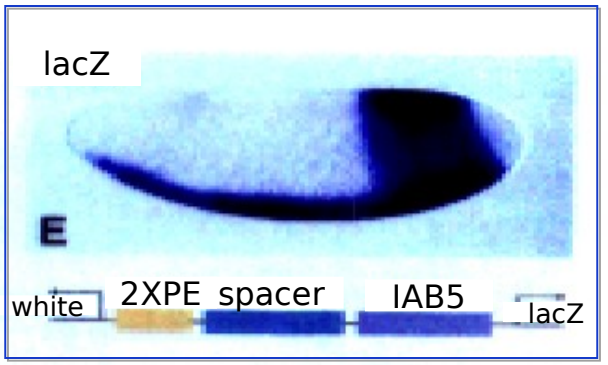
H1: anterior E — IAB5: posterior E

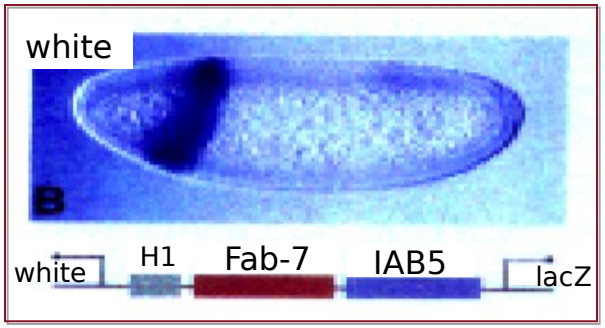
8

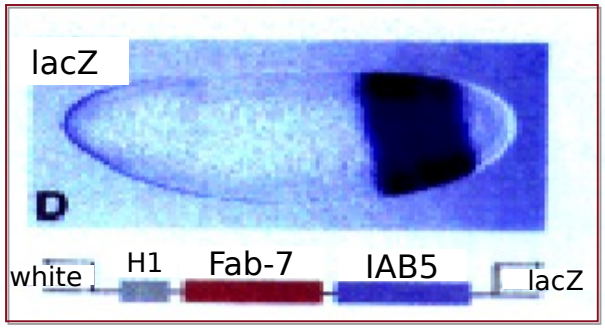
2X PE: neutral

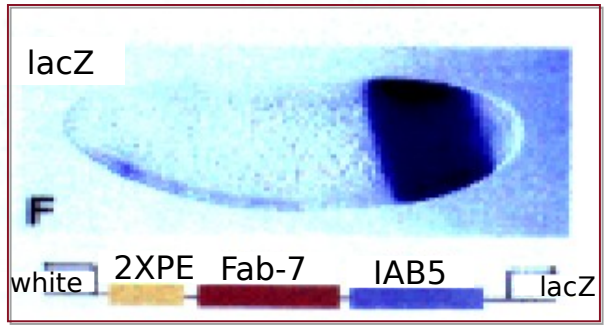


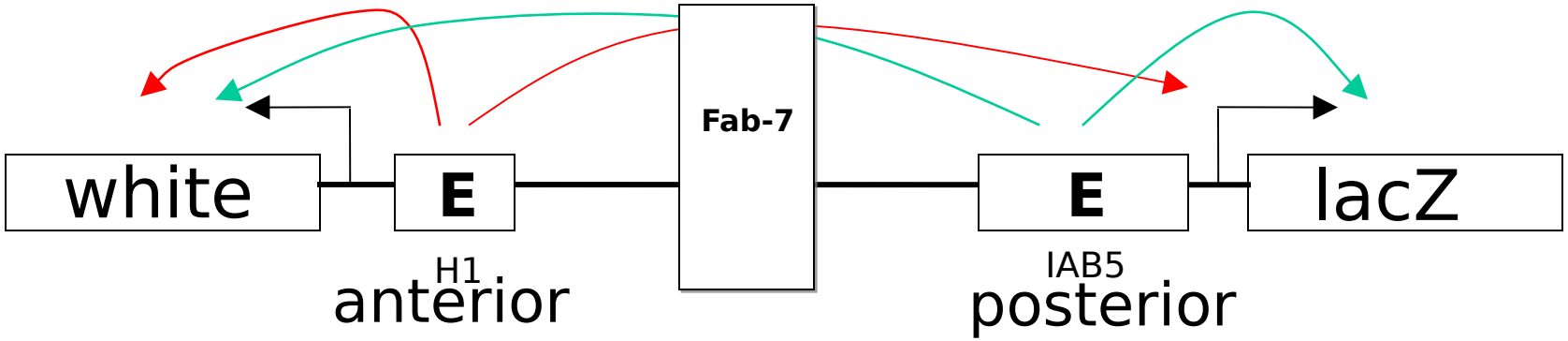






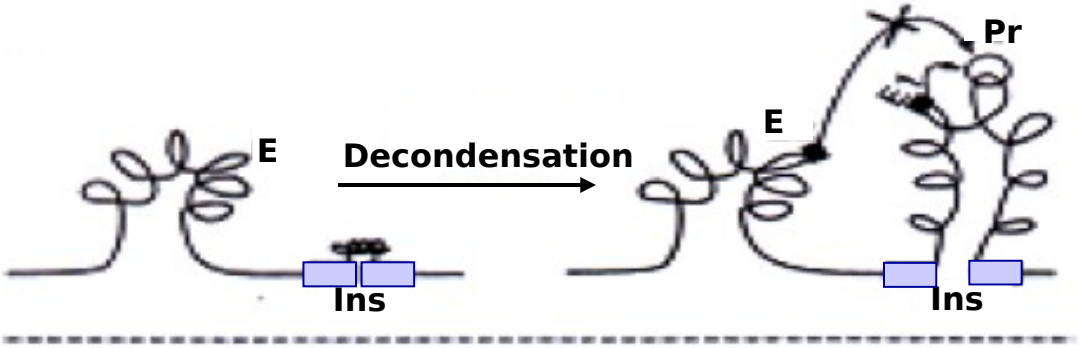




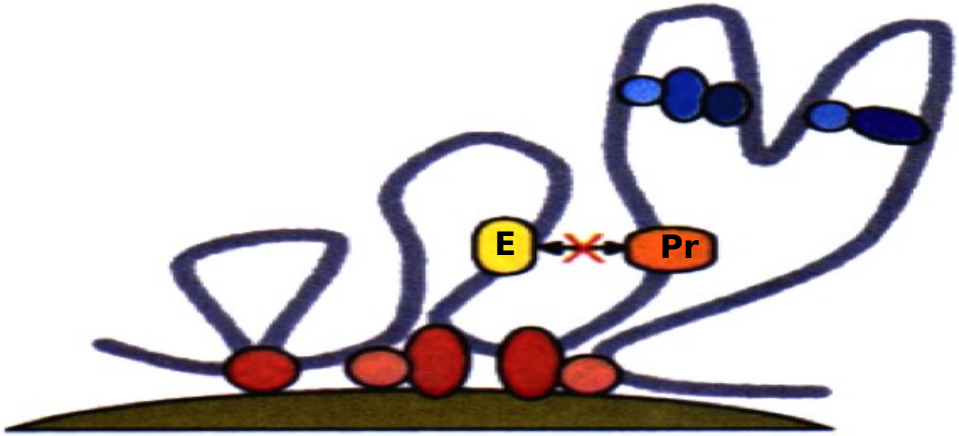


MODELS OF INSULATION

Chromatin domain



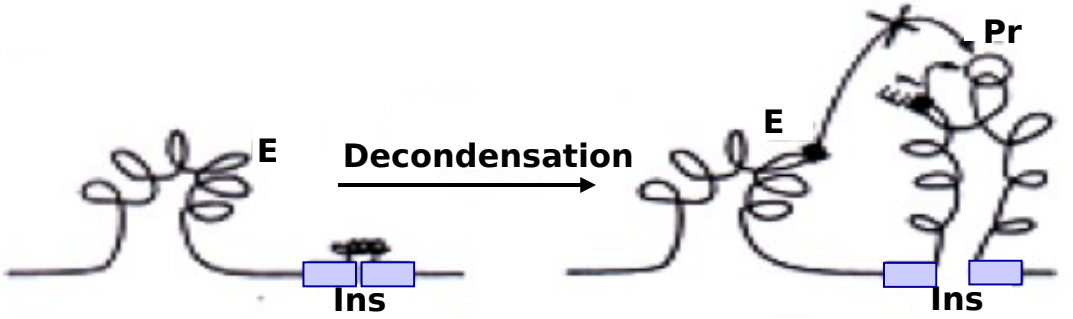
Loop Domain Model



Green: Structural Components
Red: DNA binding proteins
Yellow: Enhancer
Orange: Promoter
Blue: 'Other' proteins

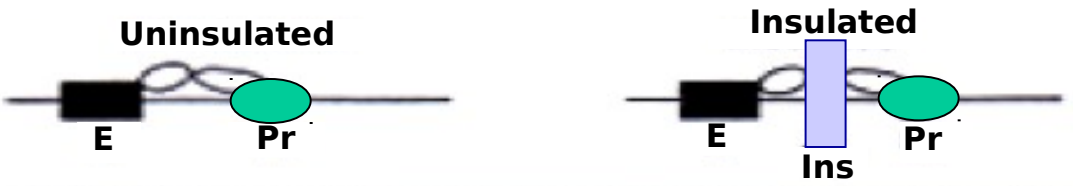
MODELS OF INSULATION

Chromatin domain

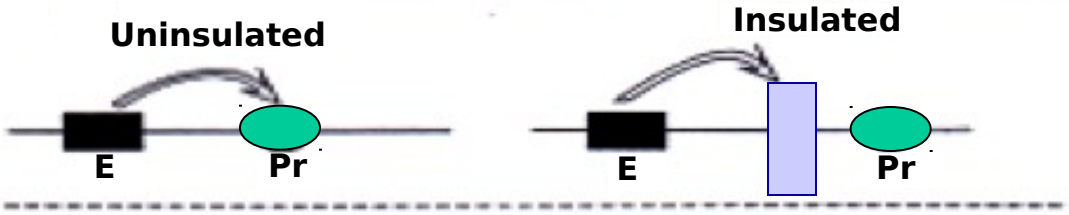


Tracking

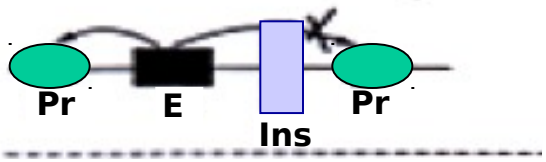
Blocking



Decoy

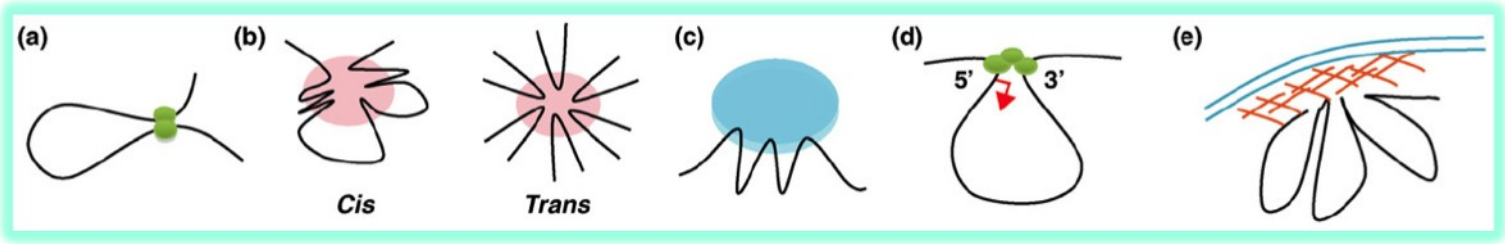


Does not kill E activity



Therefore not repressive chromatin structure!

Investigating chromatin interactions involvin



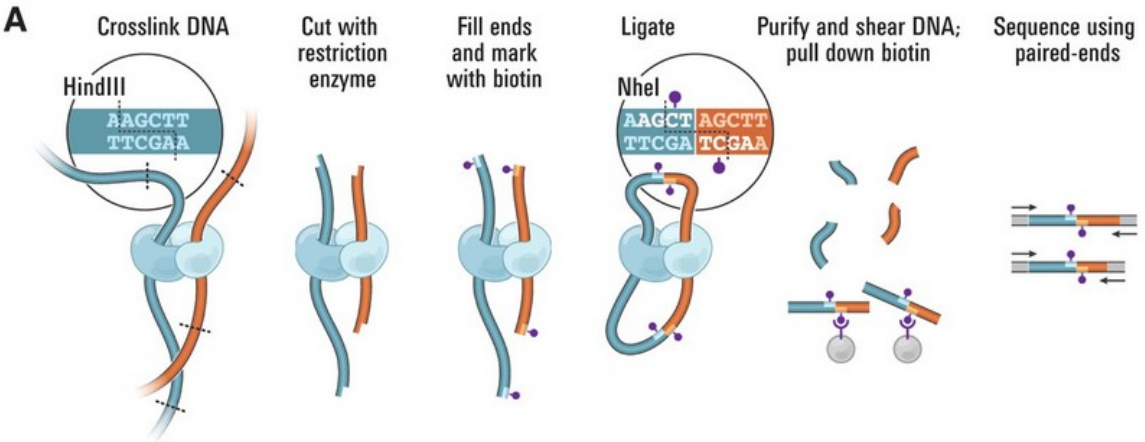
promoter/enhancer

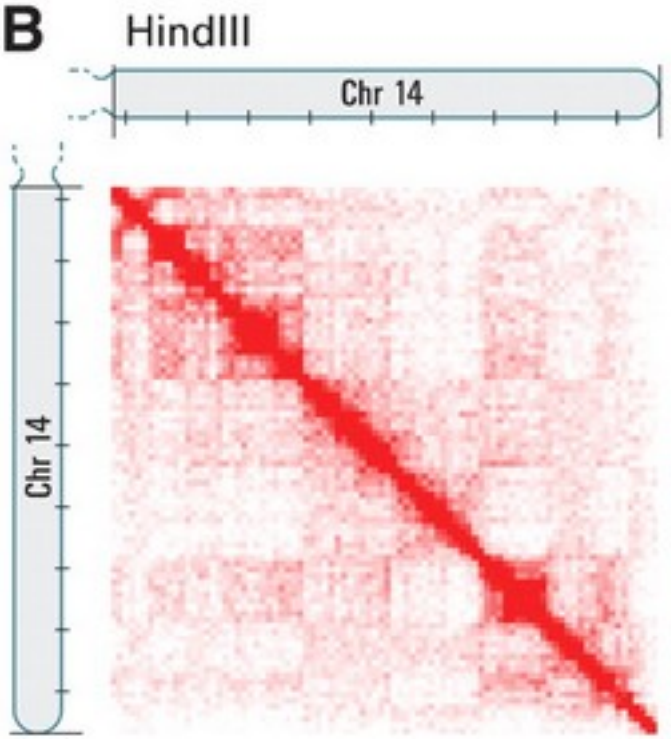
multiple regions cluster together, e.g. b-globin locus

loops result from assoc of distal elements with shared subnuclear struc-tures (e.g. insulator bodies)

promoter/ terminator

chromatin loops can be linked to targeting the nuclear periphery (e.g. nuclear lamina or pores)



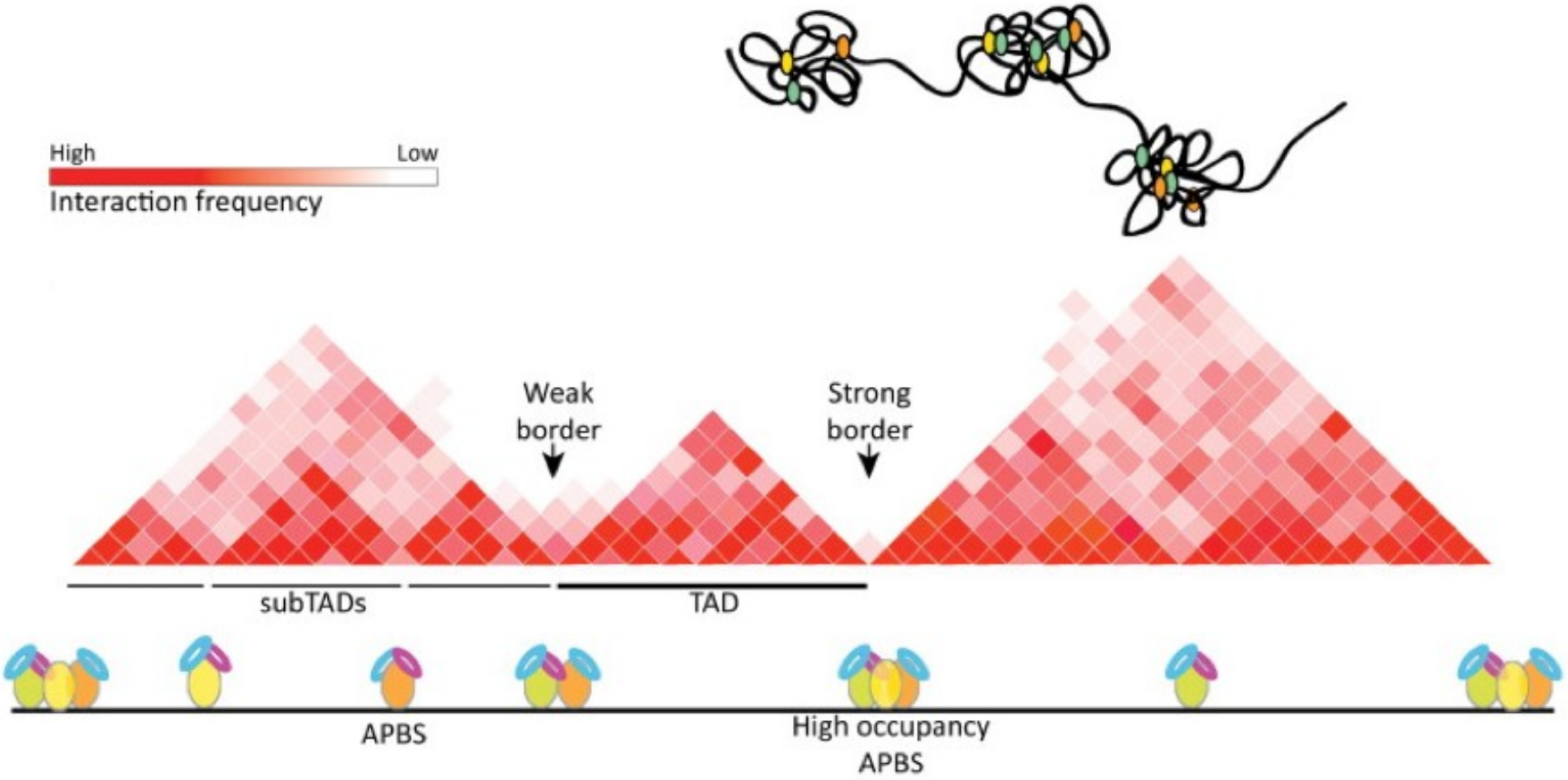


(B) Hi-C produces a genome-wide contact matrix. The submatrix shown here corresponds to intrachromosomal interactions on chromosome 14. Each pixel represents all interactions between a 1Mb locus and another 1Mb locus; intensity corresponds to the total number of reads.

The topological domains comes later which are basically regions of genome where the elements involved in looping tends to happen in one domain.

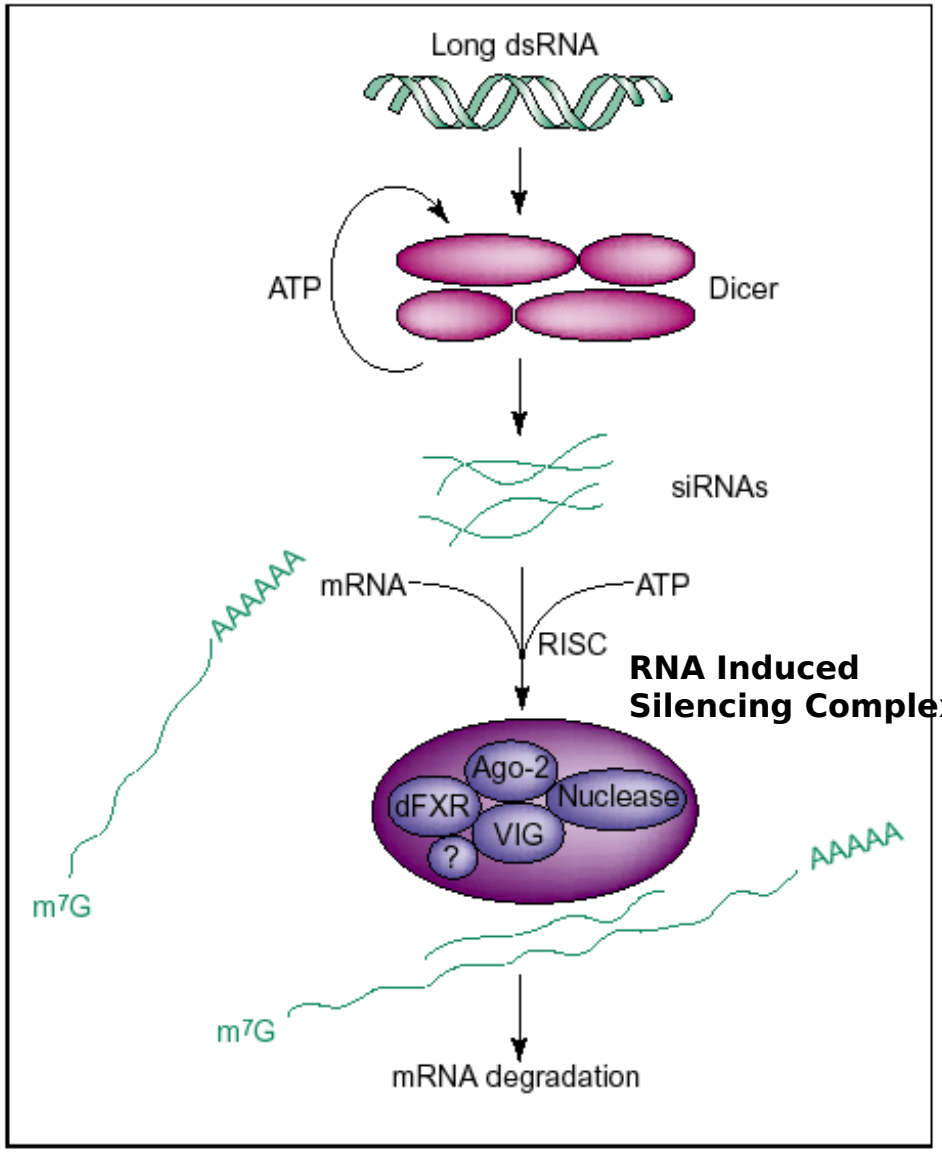
Interaction frequencies reveal ‘topologically associating domai

TADs are defined as regions of the genome undergoing high frequency of local interactions. They are separated by borders that preclude interactions between adjacent TADs. Highly occupied architectural protein binding sites (APBSs), containing multiple architectural proteins, are enriched at TAD borders, whereas low-occupancy APBSs are enriched inside TADs.

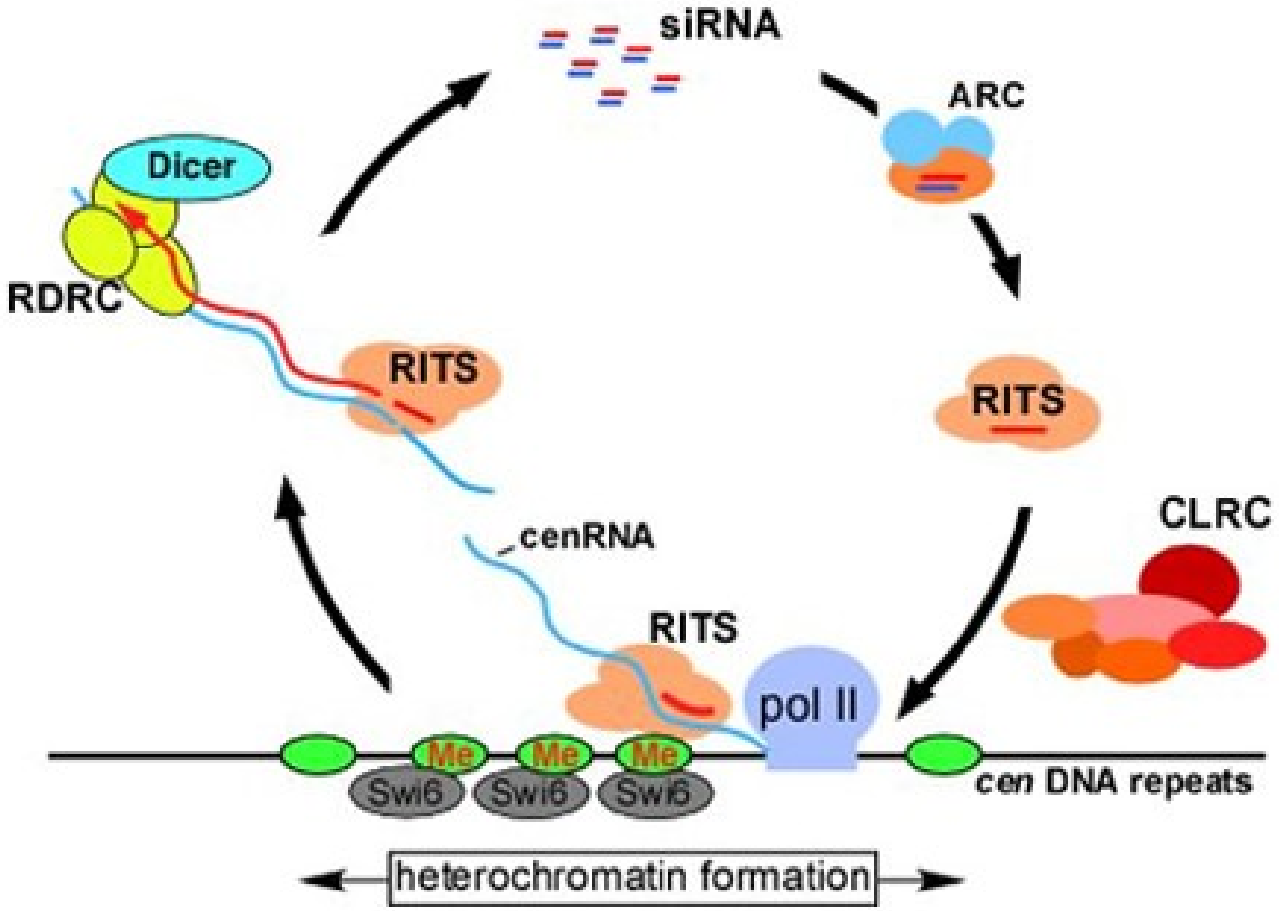


RNAi

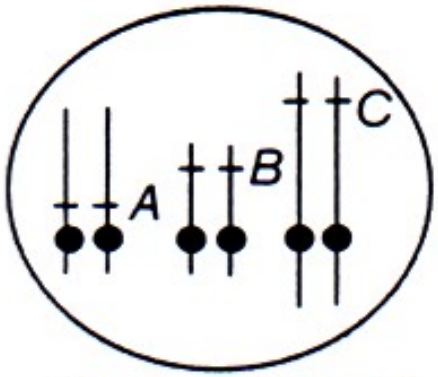
RNAi mediated mRNA degradation



RNAi also mediates the formation of centromeric heterochro



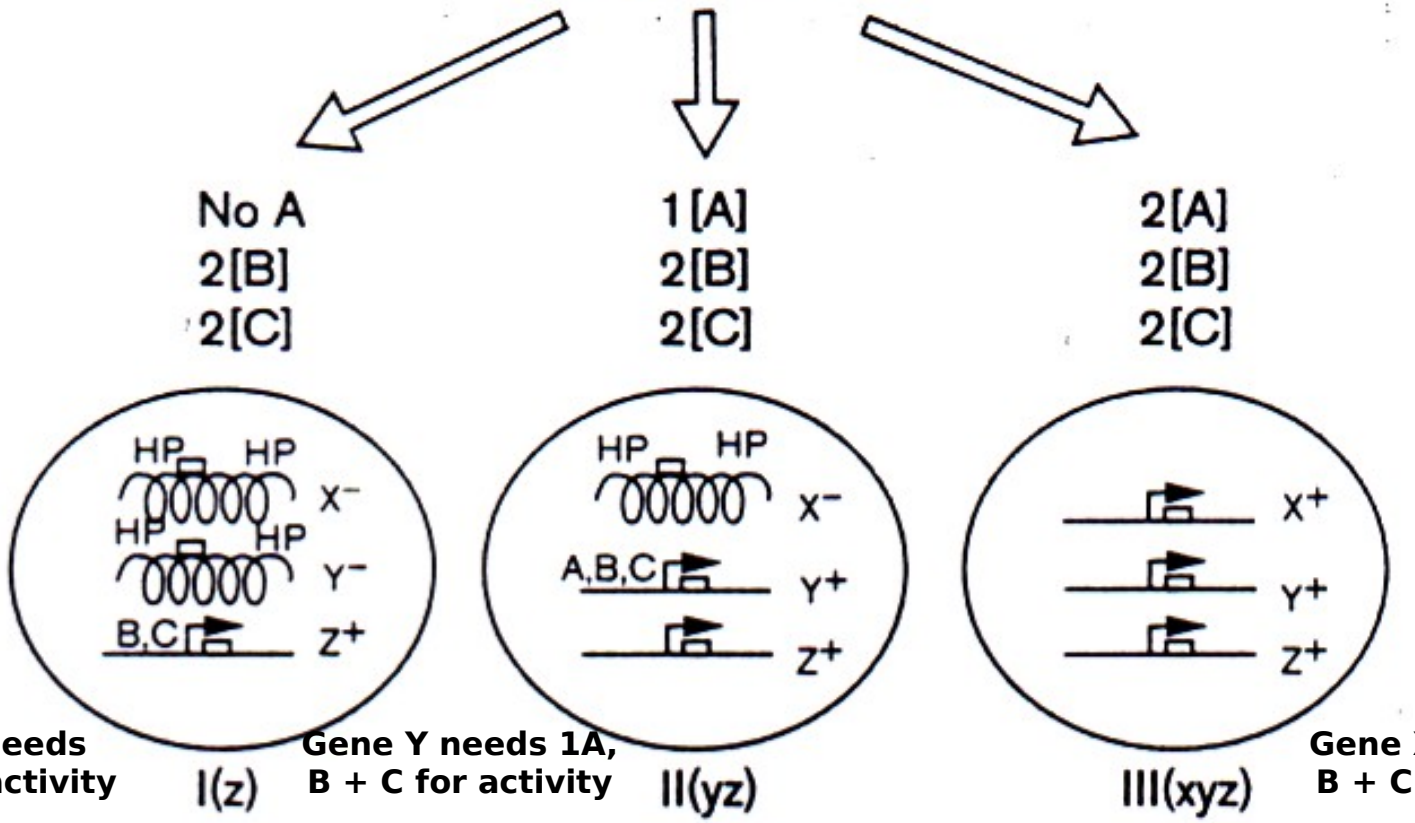
Gene A is near the centromere and it has no LCR - subject to PEV i.e. subject to stochastic decisions



Genes B + C are constitutively expressed

A, B, and C encode transcripti factors

Population of progenitor cells



Gene Z needs B + C for activity

I(z)

Gene Y needs 1A, B + C for activity

II(yz)

III(xyz)

Gene X needs 2A, B + C for activity

Lineage-committed daughter cells

Reading:

1. Genomics of long-range regulatory elements
   Noonan J.P., McCallion A.S.
   Ann. Rev. Genomics Hum. Genet. 2010, 11, 1-

2. Chromatin and epigenetic features of long-range gene regulation
   Harmston N. and Lenhard B.
   Nucleic Acids Research, 2013, 41, 7185-7199
   (doi:10.1093/nar/gkt499)

3. Comprehensive mapping of long range interactions rev principles of the human genome
   Lieberman-Aiden *et. al*.,
   Science, 2009, 326, 289-293